

UC Davis

UC Davis Previously Published Works

Title

Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor.

Permalink

<https://escholarship.org/uc/item/8mx1q8fc>

Journal

Nature, 503(7477)

ISSN

0028-0836

Authors

Ge, Xing-Yi
Li, Jia-Lu
Li, Jia-Lu
et al.

Publication Date

2013-11-01

DOI

10.1038/nature12711

Peer reviewed

Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor

Xing-Yi Ge¹*, Jia-Lu Li¹*, Xing-Lou Yang¹*, Aleksei A. Chmura², Guangjian Zhu², Jonathan H. Epstein², Jonna K. Mazet³, Ben Hu¹, Wei Zhang¹, Cheng Peng¹, Yu-Ji Zhang¹, Chu-Ming Luo¹, Bing Tan¹, Ning Wang¹, Yan Zhu¹, Gary Crameri⁴, Shu-Yi Zhang⁵, Lin-Fa Wang^{4,6}, Peter Daszak² & Zheng-Li Shi¹

The 2002–3 pandemic caused by severe acute respiratory syndrome coronavirus (SARS-CoV) was one of the most significant public health events in recent history¹. An ongoing outbreak of Middle East respiratory syndrome coronavirus² suggests that this group of viruses remains a key threat and that their distribution is wider than previously recognized. Although bats have been suggested to be the natural reservoirs of both viruses^{3–5}, attempts to isolate the progenitor virus of SARS-CoV from bats have been unsuccessful. Diverse SARS-like coronaviruses (SL-CoVs) have now been reported from bats in China, Europe and Africa^{5–8}, but none is considered a direct progenitor of SARS-CoV because of their phylogenetic disparity from this virus and the inability of their spike proteins to use the SARS-CoV cellular receptor molecule, the human angiotensin converting enzyme II (ACE2)^{9,10}. Here we report whole-genome sequences of two novel bat coronaviruses from Chinese horseshoe bats (family: Rhinolophidae) in Yunnan, China: RsSHC014 and Rs3367. These viruses are far more closely related to SARS-CoV than any previously identified bat coronaviruses, particularly in the receptor binding domain of the spike protein. Most importantly, we report the first recorded isolation of a live SL-CoV (bat SL-CoV-WIV1) from bat faecal samples in Vero E6 cells, which has typical coronavirus morphology, 99.9% sequence identity to Rs3367 and uses ACE2 from humans, civets and Chinese horseshoe bats for cell entry. Preliminary *in vitro* testing indicates that WIV1 also has a broad species tropism. Our results provide the strongest evidence to date that Chinese horseshoe bats are natural reservoirs of SARS-CoV, and that intermediate hosts may not be necessary for direct human infection by some bat SL-CoVs. They also highlight the importance of pathogen-discovery programs targeting high-risk wildlife groups in emerging disease hotspots as a strategy for pandemic preparedness.

The 2002–3 pandemic of SARS¹ and the ongoing emergence of the Middle East respiratory syndrome coronavirus (MERS-CoV)² demonstrate that CoVs are a significant public health threat. SARS-CoV was shown to use the human ACE2 molecule as its entry receptor, and this is considered a hallmark of its cross-species transmissibility¹¹. The receptor binding domain (RBD) located in the amino-terminal region (amino acids 318–510) of the SARS-CoV spike (S) protein is directly involved in binding to ACE2 (ref. 12). However, despite phylogenetic evidence that SARS-CoV evolved from bat SL-CoVs, all previously identified SL-CoVs have major sequence differences from SARS-CoV in the RBD of their S proteins, including one or two deletions^{6,9}. Replacing the RBD of one SL-CoV S protein with SARS-CoV S conferred the ability to use human ACE2 and replicate efficiently in mice^{9,13}. However, to date, no SL-CoVs have been isolated from bats, and no wild-type SL-CoV of bat origin has been shown to use ACE2.

We conducted a 12-month longitudinal survey (April 2011–September 2012) of SL-CoVs in a colony of *Rhinolophus sinicus* at a single location

in Kunming, Yunnan Province, China (Extended Data Table 1). A total of 117 anal swabs or faecal samples were collected from individual bats using a previously published method^{5,14}. A one-step reverse transcription (RT)-nested PCR was conducted to amplify the RNA-dependent RNA polymerase (RdRP) motifs A and C, which are conserved among alphacoronaviruses and betacoronaviruses¹⁵.

Twenty-seven of the 117 samples (23%) were classed as positive by PCR and subsequently confirmed by sequencing. The species origin of all positive samples was confirmed to be *R. sinicus* by cytochrome b sequence analysis, as described previously¹⁶. A higher prevalence was observed in samples collected in October (30% in 2011 and 48.7% in 2012) than those in April (7.1% in 2011) or May (7.4% in 2012) (Extended Data Table 1). Analysis of the S protein RBD sequences indicated the presence of seven different strains of SL-CoVs (Fig. 1a and Extended Data Figs 1 and 2). In addition to RBD sequences, which closely matched previously described SL-CoVs (Rs672, Rf1 and HKU3)^{5,8,17,18}, two novel strains (designated SL-CoV RsSHC014 and Rs3367) were discovered. Their full-length genome sequences were determined, and both were found to be 29,787 base pairs in size (excluding the poly(A) tail). The overall nucleotide sequence identity of these two genomes with human SARS-CoV (Tor2 strain) is 95%, higher than that observed previously for bat SL-CoVs in China (88–92%)^{5,8,17,18} or Europe (76%)⁶ (Extended Data Table 2 and Extended Data Figs 3 and 4). Higher sequence identities were observed at the protein level between these new SL-CoVs and SARS-CoVs (Extended Data Tables 3 and 4). To understand the evolutionary origin of these two novel SL-CoV strains, we conducted recombination analysis with the Recombination Detection Program 4.0 package¹⁹ using available genome sequences of bat SL-CoV strains (Rf1, Rp3, Rs672, Rm1, HKU3 and BM48-31) and human and civet representative SARS-CoV strains (BJ01, SZ3, Tor2 and GZ02). Three breakpoints were detected with strong *P* values ($<10^{-20}$) and supported by similarity plot and bootscan analysis (Extended Data Fig. 5a, b). Breakpoints were located at nucleotides 20,827, 26,553 and 28,685 in the Rs3367 (and RsSHC014) genome, and generated recombination fragments covering nucleotides 20,827–26,533 (5,727 nucleotides) (including partial open reading frame (ORF) 1b, full-length S, ORF3, E and partial M gene) and nucleotides 26,534–28,685 (2,133 nucleotides) (including partial ORF M, full-length ORF6, ORF7, ORF8 and partial N gene). Phylogenetic analysis using the major and minor parental regions suggested that Rs3367, or RsSHC014, is the descendent of a recombination of lineages that ultimately lead to SARS-CoV and SL-CoV Rs672 (Fig. 1b).

The most notable sequence differences between these two new SL-CoVs and previously identified SL-CoVs is in the RBD regions of their S proteins. First, they have higher amino acid sequence identity to SARS-CoV (85% and 96% for RsSHC014 and Rs3367, respectively). Second, there are no deletions and they have perfect sequence alignment with the SARS-CoV RBD region (Extended Data Figs 1 and 2). Structural

¹Center for Emerging Infectious Diseases, State Key Laboratory of Virology, Wuhan Institute of Virology of the Chinese Academy of Sciences, Wuhan 430071, China. ²EcoHealth Alliance, New York, New York 10001, USA. ³One Health Institute, School of Veterinary Medicine, University of California, Davis, California 95616, USA. ⁴CSIRO Australian Animal Health Laboratory, Geelong, Victoria 3220, Australia. ⁵College of Life Sciences, East China Normal University, Shanghai 200062, China. ⁶Emerging Infectious Diseases Program, Duke-NUS Graduate Medical School, Singapore 169857.

*These authors contributed equally to this work.

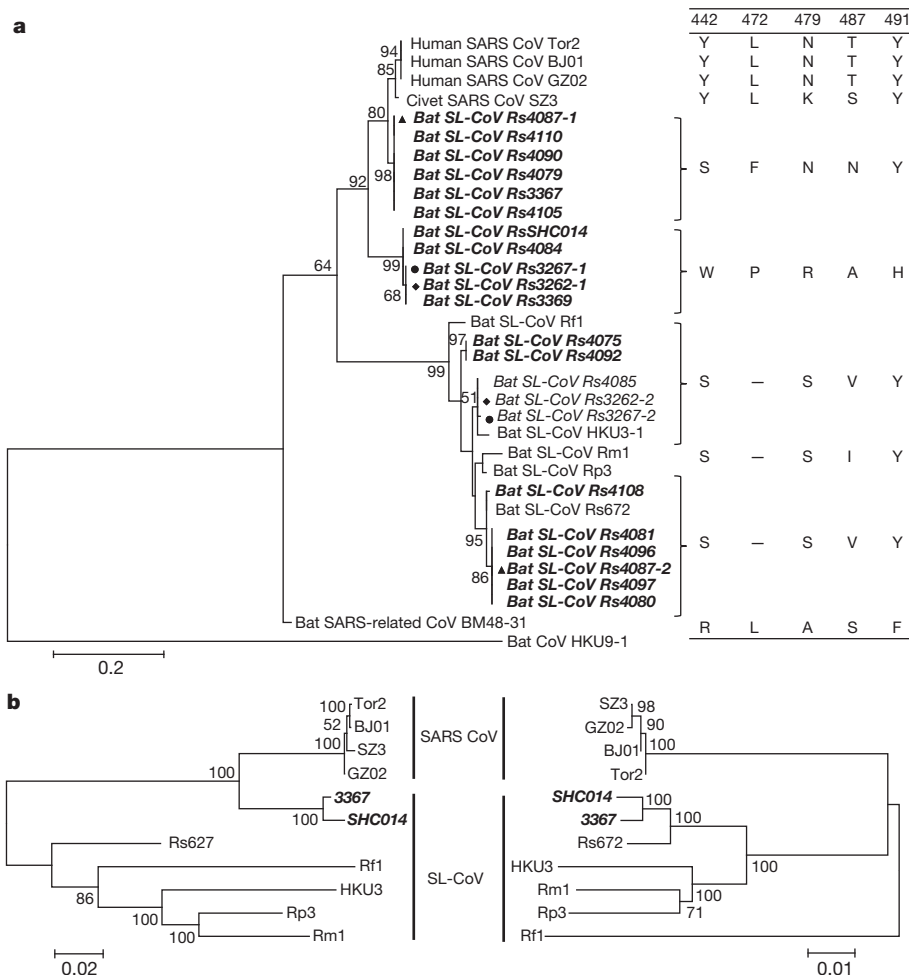


Figure 1 | Phylogenetic tree based on amino acid sequences of the S RBD region and the two parental regions of bat SL-CoV Rs3367 or RsSHC014. **a**, SARS-CoV S protein amino acid residues 310–520 were aligned with homologous regions of bat SL-CoVs using the ClustalW software. A maximum-likelihood phylogenetic tree was constructed using a Poisson model with bootstrap values determined by 1,000 replicates in the MEGA5 software package. The RBD sequences identified in this study are in bold and named by the sample numbers. The key amino acid residues involved in interacting with the human ACE2 molecule are indicated on the right of the tree. SARS-CoV GZ02, BJ01 and Tor2 were isolated from patients in the early, middle and late phase, respectively, of the SARS outbreak in 2003. SARS-CoV SZ3 was identified from *Paguma larvata* in 2003 collected in Guangdong, China. SL-CoV Rp3, Rs672 and HKU3-1 were identified from *R. sinicus* collected in China (respectively: Guangxi, 2004; Guizhou, 2006; Hong Kong, 2005). Rf1 and Rm1 were identified from

R. ferrumequinum and *R. macrotis*, respectively, collected in Hubei, China, in 2004. Bat SARS-related CoV BM48-31 was identified from *R. blasii* collected in Bulgaria in 2008. Bat CoV HKU9-1 was identified from *Rousettus leschenaultii* collected in Guangdong, China in 2005/2006 and used as an outgroup. All sequences in bold and italics were identified in the current study. Filled triangles, circles and diamonds indicate samples with co-infection by two different SL-CoVs. “-” indicates the amino acid deletion. **b**, Phylogenetic origins of the two parental regions of Rs3367 or RsSHC014. Maximum likelihood phylogenetic trees were constructed from alignments of two fragments covering nucleotides 20,827–26,533 (5,727 nucleotides) and 26,534–28,685 (2,133 nucleotides) of the Rs3367 genome, respectively. For display purposes, the trees were midpoint rooted. The taxa were annotated according to strain names: SARS-CoV, SARS coronavirus; SARS-like CoV, bat SARS-like coronavirus. The two novel SL-CoVs, Rs3367 and RsSHC014, are in bold and italics.

and mutagenesis studies have previously identified five key residues (amino acids 442, 472, 479, 487 and 491) in the RBD of the SARS-CoV S protein that have a pivotal role in receptor binding^{20,21}. Although all five residues in the RsSHC014 S protein were found to be different from those of SARS-CoV, two of the five residues in the Rs3367 RBD were conserved (Fig. 1 and Extended Data Fig. 1).

Despite the rapid accumulation of bat CoV sequences in the last decade, there has been no report of successful virus isolation^{6,22,23}. We attempted isolation from SL-CoV PCR-positive samples. Using an optimized protocol and Vero E6 cells, we obtained one isolate which caused cytopathic effect during the second blind passage. Purified virions displayed typical coronavirus morphology under electron microscopy (Fig. 2). Sequence analysis using a sequence-independent amplification method¹⁴ to avoid PCR-introduced contamination indicated that the isolate was almost identical to Rs3367, with 99.9% nucleotide genome sequence identity and 100% amino acid sequence identity for the S1 region. The new isolate was named SL-CoV-WIV1.

To determine whether WIV1 can use ACE2 as a cellular entry receptor, we conducted virus infectivity studies using HeLa cells expressing or not expressing ACE2 from humans, civets or Chinese horseshoe bats. We found that WIV1 is able to use ACE2 of different origins as an entry receptor and replicated efficiently in the ACE2-expressing cells (Fig. 3). This is, to our knowledge, the first identification of a wild-type bat SL-CoV capable of using ACE2 as an entry receptor.

To assess its cross-species transmission potential, we conducted infectivity assays in cell lines from a range of species. Our results (Fig. 4 and Extended Data Table 5) indicate that bat SL-CoV-WIV1 can grow in human alveolar basal epithelial (A549), pig kidney 15 (PK-15) and *Rhinolophus sinicus* kidney (RSKT) cell lines, but not in human cervix (HeLa), Syrian golden hamster kidney (BHK21), *Myotis davidii* kidney (BK), *Myotis chinensis* kidney (MCKT), *Rousettus leschenaultii* kidney (RLK) or *Pteropus alecto* kidney (PaKi) cell lines. Real-time RT-PCR indicated that WIV1 replicated much less efficiently in A549, PK-15 and RSKT cells than in Vero E6 cells (Fig. 4).

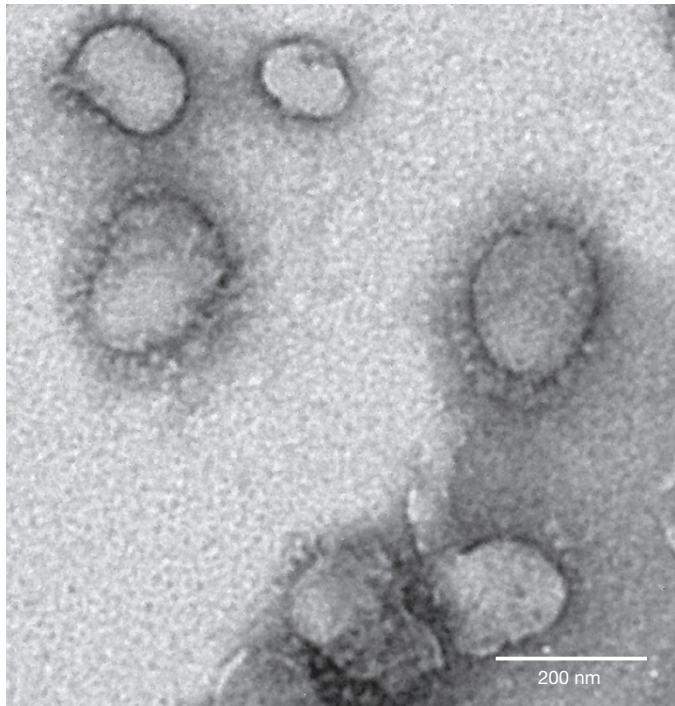


Figure 2 | Electron micrograph of purified virions. Virions from a 10-ml culture were collected, fixed and concentrated/purified by sucrose gradient centrifugation. The pelleted viral particles were suspended in 100 μ l PBS, stained with 2% phosphotungstic acid (pH 7.0) and examined directly using a Tecnai transmission electron microscope (FEI) at 200 kV.

To assess the cross-neutralization activity of human SARS-CoV sera against WIV1, we conducted serum-neutralization assays using nine convalescent sera from SARS patients collected in 2003. The results showed that seven of these were able to completely neutralize 100 tissue

culture infectious dose 50 (TCID₅₀) WIV1 at dilutions of 1:10 to 1:40, further confirming the close relationship between WIV1 and SARS-CoV.

Our findings have important implications for public health. First, they provide the clearest evidence yet that SARS-CoV originated in bats. Our previous work provided phylogenetic evidence of this⁵, but the lack of an isolate or evidence that bat SL-CoVs can naturally infect human cells, until now, had cast doubt on this hypothesis. Second, the lack of capacity of SL-CoVs to use of ACE2 receptors has previously been considered as the key barrier for their direct spillover into humans, supporting the suggestion that civets were intermediate hosts for SARS-CoV adaptation to human transmission during the SARS outbreak²⁴. However, the ability of SL-CoV-WIV1 to use human ACE2 argues against the necessity of this step for SL-CoV-WIV1 and suggests that direct bat-to-human infection is a plausible scenario for some bat SL-CoVs. This has implications for public health control measures in the face of potential spillover of a diverse and growing pool of recently discovered SARS-like CoVs with a wide geographic distribution.

Our findings suggest that the diversity of bat CoVs is substantially higher than that previously reported. In this study we were able to demonstrate the circulation of at least seven different strains of SL-CoVs within a single colony of *R. sinicus* during a 12-month period. The high genetic diversity of SL-CoVs within this colony was mirrored by high phenotypic diversity in the differential use of ACE2 by different strains. It would therefore not be surprising if further surveillance reveals a broad diversity of bat SL-CoVs that are able to use ACE2, some of which may have even closer homology to SARS-CoV than SL-CoV-WIV1. Our results—in addition to the recent demonstration of MERS-CoV in a Saudi Arabian bat²⁵, and of bat CoVs closely related to MERS-CoV in China, Africa, Europe and North America^{3,26,27}—suggest that bat coronaviruses remain a substantial global threat to public health.

Finally, this study demonstrates the public health importance of pathogen discovery programs targeting wildlife that aim to identify the 'known unknowns'—previously unknown viral strains closely related to known pathogens. These programs, focused on specific high-risk wildlife groups and hotspots of disease emergence, may be a critical part of future global strategies to predict, prepare for, and prevent pandemic emergence²⁸.

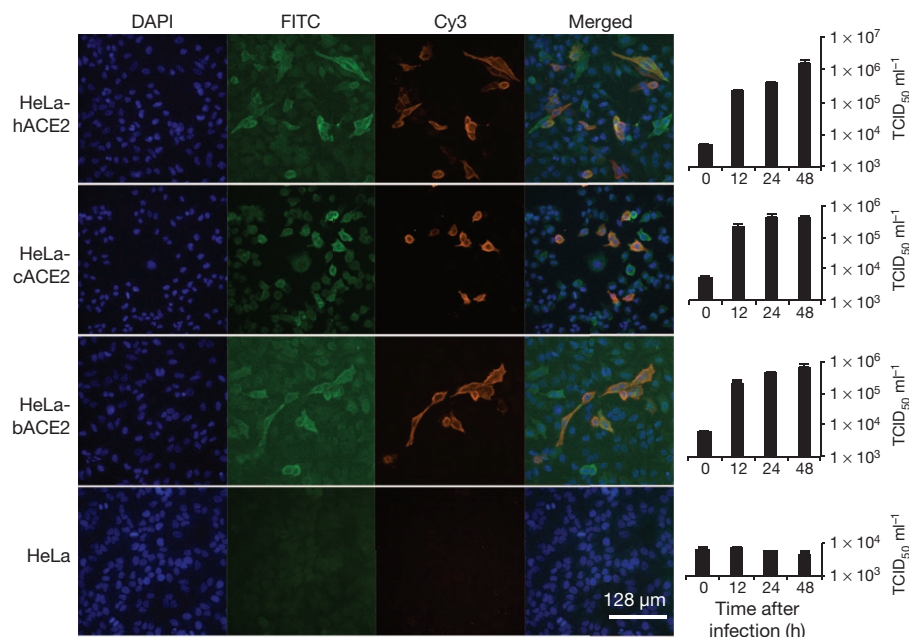


Figure 3 | Analysis of receptor usage of SL-CoV-WIV1 determined by immunofluorescence assay and real-time PCR. Determination of virus infectivity in HeLa cells with and without the expression of ACE2. b, bat; c, civet; h, human. ACE2 expression was detected with goat anti-humanACE2 antibody followed by fluorescein isothiocyanate (FITC)-conjugated donkey anti-goat IgG. Virus replication was detected with rabbit antibody against the

SL-CoV Rp3 nucleocapsid protein followed by cyanine 3 (Cy3)-conjugated mouse anti-rabbit IgG. Nuclei were stained with DAPI (4',6-diamidino-2-phenylindole). The columns (from left to right) show staining of nuclei (blue), ACE2 expression (green), virus replication (red), merged triple-stained images and real-time PCR results, respectively. ($n = 3$); error bars represent standard deviation.

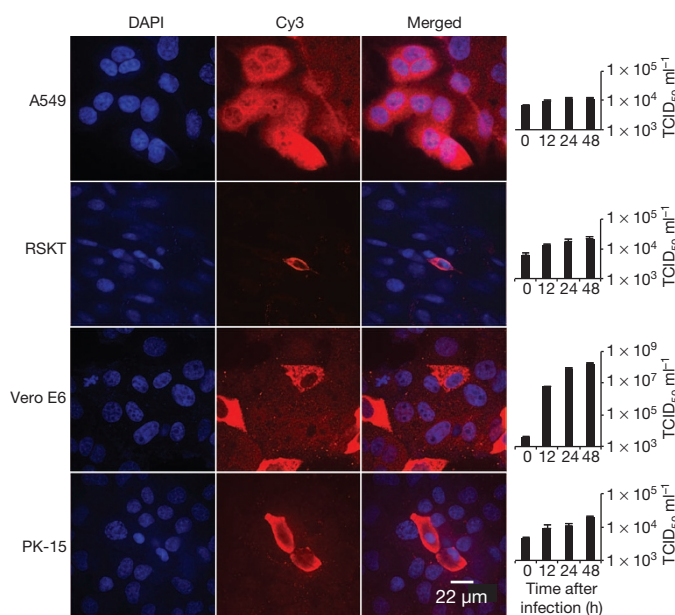


Figure 4 | Analysis of host range of SL-CoV-WIV1 determined by immunofluorescence assay and real-time PCR. Virus infection in A549, RSKT, Vero E6 and PK-15 cells. Virus replication was detected as described for Fig. 3. The columns (from left to right) show staining of nuclei (blue), virus replication (red), merged double-stained images and real-time PCR results, respectively. $n = 3$; error bars represent s.d.

METHODS SUMMARY

Throat and faecal swabs or fresh faecal samples were collected in viral transport medium as described previously¹⁴. All PCR was conducted with the One-Step RT-PCR kit (Invitrogen). Primers targeting the highly conserved regions of the RdRP gene were used for detection of all alphacoronaviruses and betacoronaviruses as described previously¹⁵. Degenerate primers were designed on the basis of all available genomic sequences of SARS-CoVs and SL-CoVs and used for amplification of the RBD sequences of S genes or full-length genomic sequences. Degenerate primers were used for amplification of the bat ACE2 gene as described previously²⁹. PCR products were gel purified and cloned into pGEM-T Easy Vector (Promega). At least four independent clones were sequenced to obtain a consensus sequence. PCR-positive faecal samples (in 200 μ l buffer) were gradient centrifuged at 3,000–12,000g and supernatant diluted at 1:10 in DMEM before being added to Vero E6 cells. After incubation at 37 °C for 1 h, inocula were removed and replaced with fresh DMEM with 2% FCS. Cells were incubated at 37 °C and checked daily for cytopathic effect. Cell lines from different origins were grown on coverslips in 24-well plates and inoculated with the novel SL-CoV at a multiplicity of infection of 10. Virus replication was detected at 24 h after infection using rabbit antibodies against the SL-CoV Rp3 nucleocapsid protein followed by Cy3-conjugated goat anti-rabbit IgG.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 May; accepted 18 September 2013.

Published online 30 October 2013.

- Ksiazek, T. G. *et al.* A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1953–1966 (2003).
- Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814–1820 (2012).
- Anthony, S. J. *et al.* Coronaviruses in bats from Mexico. *J. Gen. Virol.* **94**, 1028–1038 (2013).
- Raj, V. S. *et al.* Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* **495**, 251–254 (2013).
- Li, W. *et al.* Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676–679 (2005).
- Drexler, J. F. *et al.* Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. *J. Virol.* **84**, 11336–11349 (2010).
- Tong, S. *et al.* Detection of novel SARS-like and other coronaviruses in bats from Kenya. *Emerg. Infect. Dis.* **15**, 482–485 (2009).
- Lau, S. K. P. *et al.* Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl Acad. Sci. USA* **102**, 14040–14045 (2005).

- Ren, W. *et al.* Difference in receptor usage between severe acute respiratory syndrome (SARS) coronavirus and SARS-like coronavirus of bat origin. *J. Virol.* **82**, 1899–1907 (2008).
- Hon, C. C. *et al.* Evidence of the recombinant origin of a bat severe acute respiratory syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS coronavirus. *J. Virol.* **82**, 1819–1826 (2008).
- Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454 (2003).
- Wong, S. K., Li, W., Moore, M. J., Choe, H. & Farzan, M. A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J. Biol. Chem.* **279**, 3197–3201 (2004).
- Becker, M. M. *et al.* Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc. Natl Acad. Sci. USA* **105**, 19944–19949 (2008).
- Li, Y. *et al.* Host range, prevalence, and genetic diversity of adenoviruses in bats. *J. Virol.* **84**, 3889–3897 (2010).
- De Souza Luna, L. K. *et al.* Generic detection of coronaviruses and differentiation at the prototype strain level by reverse transcription-PCR and nonfluorescent low-density microarray. *J. Clin. Microbiol.* **45**, 1049–1052 (2007).
- Cui, J. *et al.* Evolutionary relationships between bat coronaviruses and their hosts. *Emerg. Infect. Dis.* **13**, 1526–1532 (2007).
- Yuan, J. *et al.* Intraspecies diversity of SARS-like coronaviruses in *Rhinolophus sinicus* and its implications for the origin of SARS coronaviruses in humans. *J. Gen. Virol.* **91**, 1058–1062 (2010).
- Ren, W. *et al.* Full-length genome sequences of two SARS-like coronaviruses in horseshoe bats and genetic variation analysis. *J. Gen. Virol.* **87**, 3355–3359 (2006).
- Martin, D. P. *et al.* RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**, 2462–2463 (2010).
- Wu, K., Peng, G., Wilken, M., Geraghty, R. J. & Li, F. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *J. Biol. Chem.* **287**, 8904–8911 (2012).
- Li, W. *et al.* Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* **24**, 1634–1643 (2005).
- Lau, S. K. *et al.* Ecological epidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related *Rhinolophus* bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. *J. Virol.* **84**, 2808–2819 (2010).
- Lau, S. K. *et al.* Coexistence of different genotypes in the same bat and serological characterization of *Rousettus* bat coronavirus HKU9 belonging to a novel Betacoronavirus subgroup. *J. Virol.* **84**, 11385–11394 (2010).
- Song, H. D. *et al.* Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc. Natl Acad. Sci. USA* **102**, 2430–2435 (2005).
- Memish, Z. A. *et al.* Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. *Emerg. Infect. Dis.* **19**, 11 (2013).
- Chan, J. F. *et al.* Is the discovery of the novel human betacoronavirus 2c EMC/2012 (HCoV-EMC) the beginning of another SARS-like pandemic? *J. Infect.* **65**, 477–489 (2012).
- Ithete, N. L. *et al.* Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. *Emerg. Infect. Dis.* **19**, 1697–1699 (2013).
- Morse, S. S. *et al.* Prediction and prevention of the next pandemic zoonosis. *Lancet* **380**, 1956–1965 (2012).
- Hou, Y. *et al.* Angiotensin-converting enzyme 2 (ACE2) proteins of different bat species confer variable susceptibility to SARS-CoV entry. *Arch. Virol.* **155**, 1563–1569 (2010).

Acknowledgements We acknowledge financial support from the State Key Program for Basic Research (2011CB504701 and 2010CB530100), National Natural Science Foundation of China (81290341 and 31321001), Scientific and technological basis special project (2013FY113500), CSIRO OCE Science Leaders Award, National Institute of Allergy and Infectious Diseases (NIAID) award number R01AI079231, a National Institutes of Health (NIH)/National Science Foundation (NSF) 'Ecology and Evolution of Infectious Diseases' award from the NIH Fogarty International Center (R01TW005869), an award from the NIH Fogarty International Center supported by International Influenza Funds from the Office of the Secretary of the Department of Health and Human Services (R56TW009502), and United States Agency for International Development (USAID) Emerging Pandemic Threats PREDICT. The contents are the responsibility of the authors and do not necessarily reflect the views of NIAID, NIH, NSF, USAID or the United States Government. We thank X. Che from Zhujiang Hospital, Southern Medical University, for providing human SARS patient sera.

Author Contributions Z.-L.S. and P.D. designed and coordinated the study. X.-Y.G., J.-L. and X.-L.Y. conducted majority of experiments and contributed equally to the study. A.A.C., B.H., W.Z., C.P., Y.-J.Z., C.-M.L., B.T., N.W. and Y.Z. conducted parts of the experiments and analyses. J.H.E., J.K.M. and S.-Y.Z. coordinated the field study. X.-Y.G., J.-L.L., X.-L.Y., B.T. and G.-J.Z. collected the samples. G.C. and L.-F.W. designed and supervised part of the experiments. All authors contributed to the interpretations and conclusions presented. Z.-L.S. and X.-Y.G. wrote the manuscript with significant contributions from P.D. and L.-F.W. and input from all authors.

Author Information Sequences of three bat SL-CoV genomes, bat SL-CoV RBD and *R. sinicus* ACE2 genes have been deposited in GenBank under accession numbers KC881005–KC881007 (genomes from SL-CoV RsSHC014, Rs3367 and WIV1, respectively), KC880984–KC881003 (bat SL-CoV RBD genes) and KC881004 (*R. sinicus* ACE2), respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.D. (daszak@ecohealthalliance.org) or Z.-L.S. (zlshi@wh.iov.cn).

METHODS

Sampling. Bats were trapped in their natural habitat as described previously⁵. Throat and faecal swab samples were collected in viral transport medium (VTM) composed of Hank's balanced salt solution, pH 7.4, containing BSA (1%), amphotericin (15 µg ml⁻¹), penicillin G (100 U ml⁻¹) and streptomycin (50 µg ml⁻¹). To collect fresh faecal samples, clean plastic sheets measuring 2.0 by 2.0 m were placed under known bat roosting sites at about 18:00 h each evening. Relatively fresh faecal samples were collected from sheets at approximately 05:30–06:00 the next morning and placed in VTM. Samples were transported to the laboratory and stored at –80 °C until use. All animals trapped for this study were released back to their habitat after sample collection. All sampling processes were performed by veterinarians with approval from Animal Ethics Committee of the Wuhan Institute of Virology (WIVH05210201) and EcoHealth Alliance under an inter-institutional agreement with University of California, Davis (UC Davis protocol no. 16048).

RNA extraction, PCR and sequencing. RNA was extracted from 140 µl of swab or faecal samples with a Viral RNA Mini Kit (Qiagen) following the manufacturer's instructions. RNA was eluted in 60 µl RNase-free buffer (buffer AVE, Qiagen), then aliquoted and stored at –80 °C. One-step RT-PCR (Invitrogen) was used to detect coronavirus sequences as described previously¹⁵. First round PCR was conducted in a 25-µl reaction mix containing 12.5 µl PCR 2× reaction mix buffer, 10 pmol of each primer, 2.5 mM MgSO₄, 20 U RNase inhibitor, 1 µl SuperScript III/ Platinum Taq Enzyme Mix and 5 µl RNA. Amplification of the RdRP-gene fragment was performed as follows: 50 °C for 30 min, 94 °C for 2 min, followed by 40 cycles consisting of 94 °C for 15 s, 62 °C for 15 s, 68 °C for 40 s, and a final extension of 68 °C for 5 min. Second round PCR was conducted in a 25-µl reaction mix containing 2.5 µl PCR reaction buffer, 5 pmol of each primer, 50 mM MgCl₂, 0.5 mM dNTP, 0.1 µl Platinum Taq Enzyme (Invitrogen) and 1 µl first round PCR product. The amplification of RdRP-gene fragment was performed as follows: 94 °C for 5 min followed by 35 cycles consisting of 94 °C for 30 s, 52 °C for 30 s, 72 °C for 40 s, and a final extension of 72 °C for 5 min.

To amplify the RBD region, one-step RT-PCR was performed with primers designed based on available SARS-CoV or bat SL-CoVs (first round PCR primers; F, forward; R, reverse: CoVS931F-5'-VWGADGTTGTAGRTTYCCT-3' and CoVS1909R-5'-TAARACAVCCWGCYTGWGT-3'; second PCR primers: CoVS951F-5'-TGTKAGRTTYCCTAAYATTAC-3' and CoVS1805R-5'-ACATCYTG ATANARAACAGC-3'). First-round PCR was conducted in a 25-µl reaction mix as described above except primers specific for the S gene were used. The amplification of the RBD region of the S gene was performed as follows: 50 °C for 30 min, 94 °C for 2 min, followed by 35 cycles consisting of 94 °C for 15 s, 43 °C for 15 s, 68 °C for 90 s, and a final extension of 68 °C for 5 min. Second-round PCR was conducted in a 25-µl reaction mix containing 2.5 µl PCR reaction buffer, 5 pmol of each primer, 50 mM MgCl₂, 0.5 mM dNTP, 0.1 µl Platinum Taq Enzyme (Invitrogen) and 1 µl first round PCR product. Amplification was performed as follows: 94 °C for 5 min followed by 40 cycles consisting of 94 °C for 30 s, 41 °C for 30 s, 72 °C for 60 s, and a final extension of 72 °C for 5 min.

PCR products were gel purified and cloned into pGEM-T Easy Vector (Promega). At least four independent clones were sequenced to obtain a consensus sequence for each of the amplified regions.

Sequencing full-length genomes. Degenerate coronavirus primers were designed based on all available SARS-CoV and bat SL-CoV sequences in GenBank and specific primers were designed from genome sequences generated from previous rounds of sequencing in this study (primer sequences will be provided upon request). All PCRs were conducted using the One-Step RT-PCR kit (Invitrogen). The 5' and 3' genomic ends were determined using the 5' or 3' RACE kit (Roche), respectively. PCR products were gel purified and sequenced directly or following cloning into pGEM-T Easy Vector (Promega). At least four independent clones were sequenced to obtain a consensus sequence for each of the amplified regions and each region was sequenced at least twice.

Sequence analysis and databank accession numbers. Routine sequence management and analysis was carried out using DNASTar or Geneious. Sequence alignment and editing was conducted using ClustalW, BioEdit or GeneDoc. Maximum Likelihood phylogenetic trees based on the protein sequences were constructed using a Poisson model with bootstrap values determined by 1,000 replicates in the MEGA5 software package.

Sequences obtained in this study have been deposited in GenBank as follows (accession numbers given in parenthesis): full-length genome sequence of SL-CoV RsSHC014 and Rs3367 (KC881005, KC881006); full-length sequence of WIV1 S (KC881007); RBD (KC880984–KC881003); ACE2 (KC8810040). SARS-CoV sequences used in this study: human SARS-CoV strains Tor2 (AY274119), BJ01 (AY278488), GZ02 (AY390556) and civet SARS-CoV strain SZ3 (AY304486). Bat coronavirus sequences used in this study: Rs672 (FJ588686), Rp3 (DQ071615), Rf1 (DQ412042), Rm1 (DQ412043), HKU3-1 (DQ022305), BM48-31 (NC_014470), HKU9-1 (NC_009021), HKU4 (NC_009019), HKU5 (NC_009020), HKU8 (DQ249228),

HKU2 (EF203067), BtCoV512 (NC_009657), 1A (NC_010437). Other coronavirus sequences used in this study: HCoV-229E (AF304460), HCoV-OC43 (AY391777), HCoV-NL63 (AY567487), HKU1 (NC_006577), EMC (JX869059), FIPV (NC_002306), PRCV (DQ811787), BwCoV (NC_010646), MHV (AY700211), IBV (AY851295).

Amplification, cloning and expression of the bat ACE2 gene. Construction of expression clones for human and civet ACE2 in pCDNA3.1 has been described previously²⁹. Bat ACE2 was amplified from a *R. sinicus* (sample no. 3357). In brief, total RNA was extracted from bat rectal tissue using the RNeasy Mini Kit (Qiagen). First-strand complementary DNA was synthesized from total RNA by reverse transcription with random hexamers. Full-length bat ACE2 fragments were amplified using forward primer bAF2 and reverse primer bAR2 (ref. 29). The ACE2 gene was cloned into pCDNA3.1 with KpnI and XhoI, and verified by sequencing. Purified ACE2 plasmids were transfected to HeLa cells. After 24 h, lysates of HeLa cells expressing human, civet, or bat ACE2 were confirmed by western blot or immunofluorescence assay.

Western blot analysis. Lysates of cells or filtered supernatants containing pseudoviruses were separated by SDS-PAGE, followed by transfer to a nitrocellulose membrane (Millipore). For detection of S protein, the membrane was incubated with rabbit anti-Rp3 S fragment (amino acids 561–666) polyanitibodies (1:200), and the bound antibodies were detected by alkaline phosphatase (AP)-conjugated goat anti-rabbit IgG (1:1,000). For detection of HIV-1 p24 in supernatants, monoclonal antibody against HIV p24 (p24 Mab) was used as the primary antibody at a dilution of 1:1,000, followed by incubation with AP-conjugated goat anti-mouse IgG at the same dilution. To detect the expression of ACE2 in HeLa cells, goat antibody against the human ACE2 ectodomain (1:500) was used as the first antibody, followed by incubation with horseradish peroxidase-conjugated donkey anti-goat IgG (1:1,000).

Virus isolation. Vero E6 cell monolayers were maintained in DMEM supplemented with 10% FCS. PCR-positive samples (in 200 µl buffer) were gradient centrifuged at 3,000–12,000g, and supernatant were diluted 1:10 in DMEM before being added to Vero E6 cells. After incubation at 37 °C for 1 h, inocula were removed and replaced with fresh DMEM with 2% FCS. Cells were incubated at 37 °C for 3 days and checked daily for cytopathic effect. Double-dose triple antibiotics penicillin/streptomycin/amphotericin (Gibco) were included in all tissue culture media (penicillin 200 IU ml⁻¹, streptomycin 0.2 mg ml⁻¹, amphotericin 0.5 µg ml⁻¹). Three blind passages were carried out for each sample. After each passage, both the culture supernatant and cell pellet were examined for presence of virus by RT-PCR using primers targeting the RdRP or S gene. Virions in supernatant (10 ml) were collected and fixed using 0.1% formaldehyde for 4 h, then concentrated by ultracentrifugation through a 20% sucrose cushion (5 ml) at 80,000g for 90 min using a Ty90 rotor (Beckman). The pelleted viral particles were suspended in 100 µl PBS, stained with 2% phosphotungstic acid (pH 7.0) and examined using a Tecnai transmission electron microscope (FEI) at 200 kV.

Virus infectivity detected by immunofluorescence assay. Cell lines used for this study and their culture conditions are summarized in Extended Data Table 5. Virus titre was determined in Vero E6 cells by cytopathic effect (CPE) counts. Cell lines from different origins and HeLa cells expressing ACE2 from human, civet or Chinese horseshoe bat were grown on coverslips in 24-well plates (Corning) incubated with bat SL-CoV-WIV1 at a multiplicity of infection = 10 for 1 h. The inoculum was removed and washed twice with PBS and supplemented with medium. HeLa cells without ACE2 expression and Vero E6 cells were used as negative and positive controls, respectively. At 24 h after infection, cells were washed with PBS and fixed with 4% formaldehyde in PBS (pH 7.4) for 20 min at 4 °C. ACE2 expression was detected using goat anti-human ACE2 immunoglobulin (R&D Systems) followed by FITC-labelled donkey anti-goat immunoglobulin (PTGLab). Virus replication was detected using rabbit antibody against the SL-CoV Rp3 nucleocapsid protein followed by Cy3-conjugated mouse anti-rabbit IgG. Nuclei were stained with DAPI. Staining patterns were examined using a FV1200 confocal microscope (Olympus).

Virus infectivity detected by real-time RT-PCR. Vero E6, A549, PK15, RSKT and HeLa cells with or without expression of ACE2 of different origins were inoculated with 0.1 TCID₅₀ WIV-1 and incubated for 1 h at 37 °C. After removing the inoculum, the cells were cultured with medium containing 1% FBS. Supernatants were collected at 0, 12, 24 and 48 h. RNA from 140 µl of each supernatant was extracted with the Viral RNA Mini Kit (Qiagen) following manufacturer's instructions and eluted in 60 µl buffer AVE (Qiagen). RNA was quantified on the ABI StepOne system, with the TaqMan AgPath-ID One-Step RT-PCR Kit (Applied Biosystems) in a 25 µl reaction mix containing 4 µl RNA, 1 × RT-PCR enzyme mix, 1 × RT-PCR buffer, 40 pmol forward primer (5'-GTGGTGGTGACGGCA AATG-3'), 40 pmol reverse primer (5'-AAGTGAAGCTTCTGGGCCAG-3') and 12 pmol probe (5'-FAM-AAAGAGCTCAGCCCCAGATG-BHQ1-3'). Amplification parameters were 10 min at 50 °C, 10 min at 95 °C and 50 cycles of 15 s at 95 °C and 20 s at 60 °C. RNA dilutions from purified WIV-1 stock were used as a standard.

Serum neutralization test. SARS patient sera were inactivated at 56 °C for 30 min and then used for virus neutralization testing. Sera were diluted starting with 1:10

and then serially twofold diluted in 96-well cell plates to 1:40. Each 100 μ l serum dilution was mixed with 100 μ l viral supernatant containing 100 TCID₅₀ of WTV1 and incubated at 37 °C for 1 h. The mixture was added in triplicate wells of 96-well cell plates with plated monolayers of Vero E6 cells and further incubated at 37 °C for 2 days. Serum from a healthy blood donor was used as a negative control in each experiment. CPE was observed using an inverted microscope 2 days after inoculation. The neutralizing antibody titre was read as the highest dilution of serum which completely suppressed CPE in infected wells. The neutralization test was repeated twice.

Recombination analysis. Full-length genomic sequences of SL-CoV Rs3367 or RsSHC014 were aligned with those of selected SARS-CoVs and bat SL-CoVs using Clustal X. The aligned sequences were preliminarily scanned for recombination

events using Recombination Detection Program (RDP) 4.0 (ref. 19). The potential recombination events suggested by RDP owing to their strong *P* values ($<10^{-20}$) were investigated further by similarity plot and bootscan analyses implemented in Simplot 3.5.1. Phylogenetic origin of the major and minor parental regions of Rs3367 or RsSHC014 were constructed from the concatenated sequences of the essential ORFs of the major and minor parental regions of selected SARS-CoV and SL-CoVs. Two genome regions between three estimated breakpoints (20,827–26,553 and 26,554–28,685) were aligned independently using ClustalX and generated two alignments of 5,727 base pairs and 2,133 base pairs. The two alignments were used to construct maximum likelihood trees to better infer the fragment parents. All nucleotide numberings in this study are based on Rs3367 genome position.

Human SARS-CoV GZ02
Human SARS-CoV BJ01
Human SARS-CoV Tor2
Civet SARS-CoV SZ3
Bat SL-CoV Rs3367
Bat SL-CoV RsSHC014
Bat SL-CoV Rs3369
Bat SL-CoV Rs4075
Bat SL-CoV Rs4081
Bat SL-CoV Rs4085
Bat SL-CoV Rs4108
Bat SL-CoV Rs672
Bat SL-CoV Rf1
Bat SL-CoV Rp3
Bat SL-CoV Rm1
Bat SL-CoV HKU3-1
Bat SARS-related CoV BM48-31

[illegible]

Human SARS-CoV GZ02
Human SARS-CoV BJ01
Human SARS-CoV Tor2
Civet SARS-CoV SZ3
Bat SL-CoV Rs3367
Bat SL-CoV RsSHC014
Bat SL-CoV Rs3369
Bat SL-CoV Rs4075
Bat SL-CoV Rs4081
Bat SL-CoV Rs4085
Bat SL-CoV Rs4108
Bat SL-CoV Rs672
Bat SL-CoV Rf1
Bat SL-CoV Rp3
Bat SL-CoV Rm1
Bat SL-CoV HKU3-1
Bat SARS-related CoV BM48-31

RNI DATSTGNNYKYRYLRHGKLRPFPERDISNVFSPDGKPCPTPPALNCYWPFLNDYGFYTTTGTGWCOPYRVVVLVSFELLNAP
RNI DATSTGNNYKYRYLRHGKLRPFPERDISNVFSPDGKPCPTPPALNCYWPFLNDYGFYTTTGTGWCOPYRVVVLVSFELLNAP
RNI DATSTGNNYKYRYLRHGKLRPFPERDISNVFSPDGKPCPTPPALNCYWPFLNDYGFYTTTGTGWCOPYRVVVLVSFELLNAP
RNI DATSTGNNYKYRYLRHGKLRPFPERDISNVFSPDGKPCPTPPALNCYWPFLNDYGFYTTTGTGWCOPYRVVVLVSFELLNAP
RNI DATQTGNNYKYRSLRHGKLRPFPERDISNVFSPDGKPCPTPPALNCYWPFLNDYGFYTTTGTGWCOPYRVVVLVSFELLNAP
NSK DSSSTGNNYKYRWVRRSKLNPYERDLSDNIYSPGGQSCSAVGNCYNPFLRFYGFETTAGVGHQOPYRVVVLVSFELLNAP
NSK DSSSTGNNYKYRWVRRSKLNPYERDLSDNIYSPGGQSCSAVGNCYNPFLRFYGFETTAGVGHQOPYRVVVLVSFELLNAP
AKQD---VG---SYKYRSHRSKLRPFPERDLSSDE---NGVRTSTYDFNPNVPFLVQATRVVVLVSFELLNAP
AKQD---QG---QYKYRSSRKTKLKPFFERDLSSDE---NGVRTSTYDFNPNVPFLVQATRVVVLVSFELLNAP
AKQD---TG---YKYRSHRKTKLKPFFERDLSSDDG---NGVYTESTYDFNPNVPFLVQATRVVVLVSFELLNAP
AKQD---QG---QYKYRSSRKTKLKPFFERDLSSDE---NGVRTSTYDFYPTVPIVQATRVVVLVSFELLNAP
AKQD---QG---QYKYRSSRKTKLKPFFERDLSSDE---NGVRTSTYDFNPNVPFLVQATRVVVLVSFELLNAP
AKQD---VG---SYKYRSHRSKLRPFPERDLSSDE---NGVRTSTYDFNPNVPFLVQATRVVVLVSFELLNAP
AKQD---QG---QYKYRSHRKTKLKPFFERDLSSDE---NGVRTSTYDFYPSVPVQATRVVVLVSFELLNAP
AQQD---QG---QYKYRSYRKBLKLPFFERDLSSDE---NGVYTESTYDFYPSIPIVQATRVVVLVSFELLNAP
AKHD---TG---NYKYRSHRKTKLKPFFERDLSSDDG---NGVYTESTYDFNPNVPFLVQATRVVVLVSFELLNAP
NSLD---SSNEFYRPFHGKLRKPFFERDLSDNIYSPGGTCSAEGINCYPKPLASVGETOSSGTGOPYRVVVLVSFELLNAP

Human SARS-CoV GZ02
Human SARS-CoV BJ01
Human SARS-CoV Tor2
Civet SARS-CoV SZ3
Bat SL-CoV Rs3367
Bat SL-CoV RsSHC014
Bat SL-CoV Rs3369
Bat SL-CoV Rs4075
Bat SL-CoV Rs4081
Bat SL-CoV Rs4085
Bat SL-CoV Rs4108
Bat SL-CoV Rs672
Bat SL-CoV Rf1
Bat SL-CoV Rp3
Bat SL-CoV Rm1
Bat SL-CoV HKU3-1
Bat SARS-related CoV RM48-31

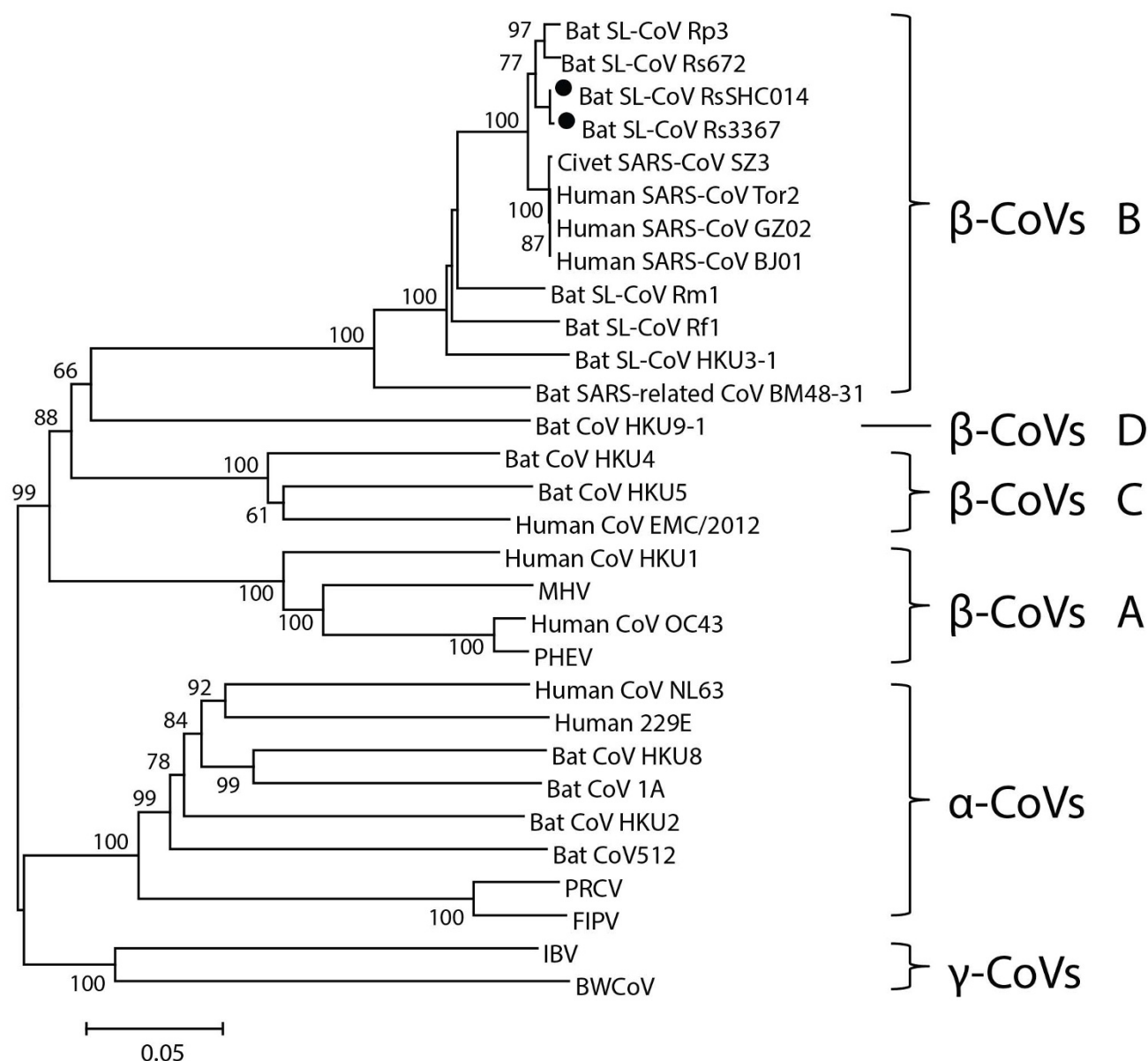
A	T	V	C	G	P	K	S	T	D	L	I	K	N	O	C	V	N	F	N	F	N	G	L	T	G	C	G	V	L	T	S	S	K	R	F	O		F	O	O	F	G	R	D	V	S	D	F	T	D	S	V	R	D	P	K	T	E	I	L	D	I	S	P	C	S	F	G	G	V	S	V	I	T	P	G	T	N
A	T	V	C	G	P	K	S	T	D	L	I	K	N	O	C	V	N	F	N	F	N	G	L	T	G	C	G	V	L	T	S	S	K	R	F	O		F	O	O	F	G	R	D	V	S	D	F	T	D	S	V	R	D	P	K	T	E	I	L	D	I	S	P	C	S	F	G	G	V	S	V	I	T	P	G	T	N
A	T	V	C	G	P	K	S	T	D	L	I	K	N	O	C	V	N	F	N	F	N	G	L	T	G	C	G	V	L	T	S	S	K	R	F	O		F	O	O	F	G	R	D	V	S	D	F	T	D	S	V	R	D	P	K	T	E	I	L	D	I	S	P	C	S	F	G	G	V	S	V	I	T	P	G	T	N
A	T	V	C	G	P	K	S	T	D	L	I	K	N	O	C	V	N	F	N	F	N	G	L	T	G	C	G	V	L	T	S	S	K	R	F	O		F	O	O	F	G	R	D	V	S	D	F	T	D	S	V	R	D	P	K	T	E	I	L	D	I	S	P	C	S	F	G	G	V	S	V	I	T	P	G	T	N
A	T	V	C	G	P	K	S	T	D	L	I	K	N	O	C	V	N	F	N	F	N	G	L	T	G	C	G	V	L	T	S	S	K	R	F	O		F	O	O	F	G	R	D	V	S	D	F	T	D	S	V	R	D	P	K	T	E	I	L	D	I	S	P	C	S	F	G	G	V	S	V	I	T	P	G	T	N
A	T	V	C	G	P	K	S	T	D	L	I	K	N	O	C	V	N	F	N	F	N	G	L	T	G	C	G	V	L	T	S	S	K	R	F	O		F	O	O	F	G	R	D	V	S	D	F	T	D	S	V	R	D	P	K	T	E	I	L	D	I	S	P	C	S	F	G	G	V	S	V	I	T	P	G	T	N
A	T	V	C	G	P	K	S	T	D	L	I	K	N	O	C	V	N	F	N	F	N	G	L	T	G	C	G	V	L	T	S	S	K	R	F	O		F	O	O	F	G	R	D	V	S	D	F	T	D	S	V	R	D	P	K	T	E	I	L	D	I	S	P	C	S	F	G	G	V	S	V	I	T	P	G	T	N
A	T	V	C	G	P	K	S	T	D	L	I	K	N	O	C	V	N	F	N	F	N	G	L	T	G	C	G	V	L	T	S	S	K	R	F	O		F	O	O	F	G	R	D	V	S	D	F	T	D	S	V	R	D	P	K	T	E	I	L	D	I	S	P	C	S	F	G	G	V	S	V	I	T	P	G	T	N
A	T	V	C	G	P	K	S	T	D	L	I	K	N	O	C	V	N	F	N	F	N	G	L	T	G	C	G	V	L	T	S	S	K	R	F	O		F	O	O	F	G	R	D	V	S	D	F	T	D	S	V	R	D	P	K	T	E	I	L	D	I	S	P	C	S	F	G	G	V	S	V	I	T	P	G	T	N
A	T	V	C	G	P	K	S	T	D	L	I	K	N	O	C	V	N	F	N	F	N	G	L	T	G	C	G	V	L	T	S	S	K	R	F	O		F	O	O	F	G	R	D	V	S	D	F	T	D	S	V	R	D	P	K	T	E	I	L	D	I	S	P	C	S	F	G	G	V	S	V	I	T	P	G	T	N
A	T	V	C	G	P	K	S	T	D	L	I	K	N	O	C	V	N	F	N</																																																											

Extended Data Figure 1 | Sequence alignment of CoV S protein RBD.

SARS-CoV S protein (amino acids 310–520) is aligned with homologous regions of bat SL-CoVs using ClustalW. The newly discovered bat SL-CoVs are

indicated with a bold vertical line on the left. The key amino acid residues involved in the interaction with human ACE2 are numbered on the top of the aligned sequences.

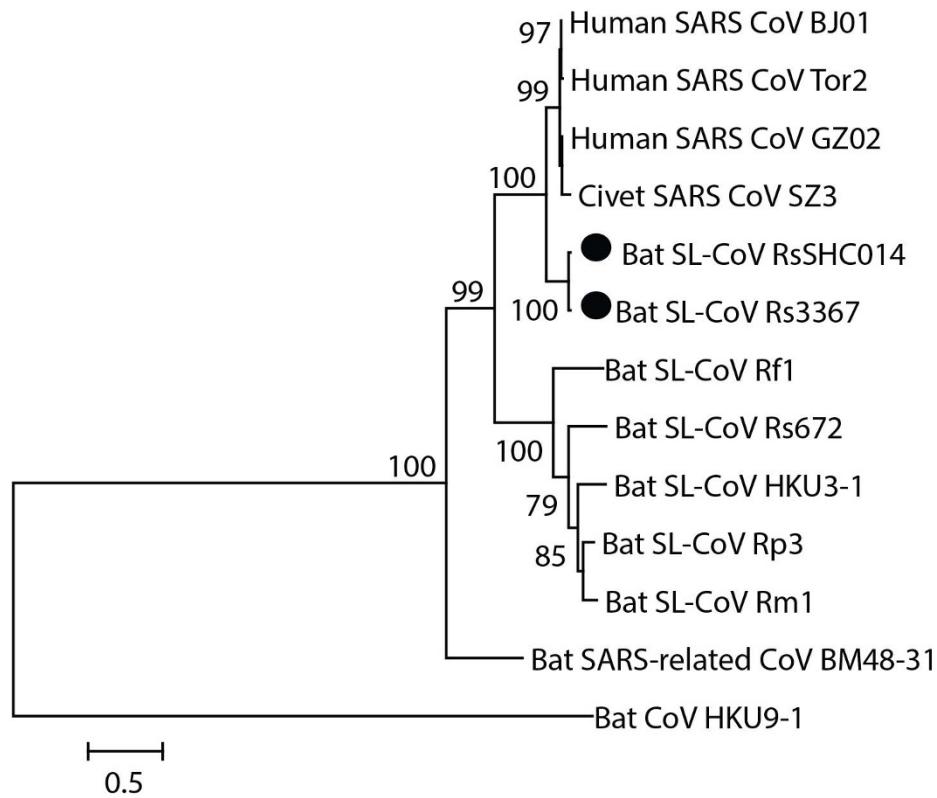
Human SARS-CoV GZ02	MFHFLPLTTSQSDLR	CTTFDDVQAENY	QHSSRGVYYP	DEIFRSD	LYLTQD	FLPP	SNV	Q	FIN	-----HT	PDNFW	PR	GDGYFAA	TEKSNV	RGWV	RG																																																																									
Human SARS-CoV BJ01	MFHFLPLTTSQSDLR	CTTFDDVQAENY	QHSSRGVYYP	DEIFRSD	LYLTQD	FLPP	SNV	Q	FIN	-----HT	PDNFW	PR	GDGYFAA	TEKSNV	RGWV	RG																																																																									
Human SARS-CoV Tor2	MFHFLPLTTSQSDLR	CTTFDDVQAENY	QHSSRGVYYP	DEIFRSD	LYLTQD	FLPP	SNV	Q	FIN	-----HT	PDNFW	PR	GDGYFAA	TEKSNV	RGWV	RG																																																																									
Civet SARS-CoV SZ3	MFHFLPLTTSQSDLR	CTTFDDVQAENY	QHSSRGVYYP	DEIFRSD	LYLTQD	FLPP	SNV	Q	FIN	-----HT	PDNFW	PR	GDGYFAA	TEKSNV	RGWV	RG																																																																									
Bat SL-CoV Rs3367	MKLEIVMPATVSSYTIK	CDPDDRTFANQ	PLSSRRGVYYP	DDIFRSD	LYLTQD	FLPP	SNV	Q	FIN	-----LN	PDNFW	PR	GDGYFAA	TEKSNV	RGWV	RG																																																																									
Bat SL-CoV RsSHC014	MKLEIVMPATVSSYTIK	CDPDDRTFANQ	PLSSRRGVYYP	DDIFRSD	LYLTQD	FLPP	SNV	Q	FIN	-----LN	PDNFW	PR	GDGYFAA	TEKSNV	RGWV	RG																																																																									
Bat SL-CoV Rs672	MKLEIVMPATVSSYTIK	CDPDDRTFANQ	PLSSRRGVYYP	DDIFRSD	LYLTQD	FLPP	SNV	Q	FIN	-----LN	PDNFW	PR	GDGYFAA	TEKSNV	RGWV	RG																																																																									
Bat SL-CoV Rf1	MKLEIVMPATVSSYTIK	CDPDDRTFANQ	PLSSRRGVYYP	DDIFRSD	LYLTQD	FLPP	SNV	Q	FIN	-----LN	PDNFW	PR	GDGYFAA	TEKSNV	RGWV	RG																																																																									
Bat SL-CoV Rm1	MKLEIVMPATVSSYTIK	CDPDDRTFANQ	PLSSRRGVYYP	DDIFRSD	LYLTQD	FLPP	SNV	Q	FIN	-----LN	PDNFW	PR	GDGYFAA	TEKSNV	RGWV	RG																																																																									
Bat SL-CoV Rp3	MKLEIVMPATVSSYTIK	CDPDDRTFANQ	PLSSRRGVYYP	DDIFRSD	LYLTQD	FLPP	SNV	Q	FIN	-----LN	PDNFW	PR	GDGYFAA	TEKSNV	RGWV	RG																																																																									
Bat SL-CoV HKU3-1	MKLEIVMPATVSSYTIK	CDPDDRTFANQ	PLSSRRGVYYP	DDIFRSD	LYLTQD	FLPP	SNV	Q	FIN	-----LN	PDNFW	PR	GDGYFAA	TEKSNV	RGWV	RG																																																																									
Bat SARS-related CoV BM48-31	MKLEIVMPATVSSYTIK	CDPDDRTFANQ	PLSSRRGVYYP	DDIFRSD	LYLTQD	FLPP	SNV	Q	FIN	-----LN	PDNFW	PR	GDGYFAA	TEKSNV	RGWV	RG																																																																									
Human SARS-CoV GZ02	STNNK	SOVI	INN	SNV	VIRA	CH	BC	ND	CF	FA	W	SK	PMG	TG	TH	FM	NA	NC	TF	EV	SD	AF	LD	VS	RG	GN	FR	HR	ER	V	FN	K	D	G	F	W	V	G	Q	I	H	D	V	R	L	D	P	S	E	N	I																																						
Human SARS-CoV BJ01	STNNK	SOVI	INN	SNV	VIRA	CH	BC	ND	CF	FA	W	SK	PMG	TG	TH	FM	NA	NC	TF	EV	SD	AF	LD	VS	RG	GN	FR	HR	ER	V	FN	K	D	G	F	W	V	G	Q	I	H	D	V	R	L	D	P	S	E	N	I																																						
Human SARS-CoV Tor2	STNNK	SOVI	INN	SNV	VIRA	CH	BC	ND	CF	FA	W	SK	PMG	TG	TH	FM	NA	NC	TF	EV	SD	AF	LD	VS	RG	GN	FR	HR	ER	V	FN	K	D	G	F	W	V	G	Q	I	H	D	V	R	L	D	P	S	E	N	I																																						
Civet SARS-CoV SZ3	STNNK	SOVI	INN	SNV	VIRA	CH	BC	ND	CF	FA	W	SK	PMG	TG	TH	FM	NA	NC	TF	EV	SD	AF	LD	VS	RG	GN	FR	HR	ER	V	FN	K	D	G	F	W	V	G	Q	I	H	D	V	R	L	D	P	S	E	N	I																																						
Bat SL-CoV Rs3367	STNNK	SOVI	INN	SNV	VIRA	CH	BC	ND	CF	FA	W	SK	PMG	TG	TH	FM	NA	NC	TF	EV	SD	AF	LD	VS	RG	GN	FR	HR	ER	V	FN	K	D	G	F	W	V	G	Q	I	H	D	V	R	L	D	P	S	E	N	I																																						
Bat SL-CoV RsSHC014	STNNK	SOVI	INN	SNV	VIRA	CH	BC	ND	CF	FA	W	SK	PMG	TG	TH	FM	NA	NC	TF	EV	SD	AF	LD	VS	RG	GN	FR	HR	ER	V	FN	K	D	G	F	W	V	G	Q	I	H	D	V	R	L	D	P	S	E	N	I																																						
Bat SL-CoV Rs672	STNNK	SOVI	INN	SNV	VIRA	CH	BC	ND	CF	FA	W	SK	PMG	TG	TH	FM	NA	NC	TF	EV	SD	AF	LD	VS	RG	GN	FR	HR	ER	V	FN	K	D	G	F	W	V	G	Q	I	H	D	V	R	L	D	P	S	E	N	I																																						
Bat SL-CoV Rf1	STNNK	SOVI	INN	SNV	VIRA	CH	BC	ND	CF	FA	W	SK	PMG	TG	TH	FM	NA	NC	TF	EV	SD	AF	LD	VS	RG	GN	FR	HR	ER	V	FN	K	D	G	F	W	V	G	Q	I	H	D	V	R	L	D	P	S	E	N	I																																						
Bat SL-CoV Rm1	STNNK	SOVI	INN	SNV	VIRA	CH	BC	ND	CF	FA	W	SK	PMG	TG	TH	FM	NA	NC	TF	EV	SD	AF	LD	VS	RG	GN	FR	HR	ER	V	FN	K	D	G	F	W	V	G	Q	I	H	D	V	R	L	D	P	S	E	N	I																																						
Bat SL-CoV Rp3	STNNK	SOVI	INN	SNV	VIRA	CH	BC	ND	CF	FA	W	SK	PMG	TG	TH	FM	NA	NC	TF	EV	SD	AF	LD	VS	RG	GN	FR	HR	ER	V	FN	K	D	G	F	W	V	G	Q	I	H	D	V	R	L	D	P	S	E	N	I																																						
Bat SL-CoV HKU3-1	STNNK	SOVI	INN	SNV	VIRA	CH	BC	ND	CF	FA	W	SK	PMG	TG	TH	FM	NA	NC	TF	EV	SD	AF	LD	VS	RG	GN	FR	HR	ER	V	FN	K	D	G	F	W	V	G	Q	I	H	D	V	R	L	D	P	S	E	N	I																																						
Bat SARS-related CoV BM48-31	STNNK	SOVI	INN	SNV	VIRA	CH	BC	ND	CF	FA	W	SK	PMG	TG	TH	FM	NA	NC	TF	EV	SD	AF	LD	VS	RG	GN	FR	HR	ER	V	FN	K	D	G	F	W	V	G	Q	I	H	D	V	R	L	D	P	S	E	N	I																																						
Human SARS-CoV GZ02	KPLKLE	IN	IT	TER	LA	TA	PA	Q	D	TS	AA	AY	FV	G	L	K	EM	IR	D	ENG	T	DA	V	DC	SN	PL	AE	LK	CS	NR	SE	LD	K	I	Y	O	T	S	N	FR	V	P	S	D	V	R	F	P	N	I	T	N	C	P	F	E	V																																
Human SARS-CoV BJ01	KPLKLE	IN	IT	TER	LA	TA	PA	Q	D	TS	AA	AY	FV	G	L	K	EM	IR	D	ENG	T	DA	V	DC	SN	PL	AE	LK	CS	NR	SE	LD	K	I	Y	O	T	S	N	FR	V	P	S	D	V	R	F	P	N	I	T	N	C	P	F	E	V																																
Human SARS-CoV Tor2	KPLKLE	IN	IT	TER	LA	TA	PA	Q	D	TS	AA	AY	FV	G	L	K	EM	IR	D	ENG	T	DA	V	DC	SN	PL	AE	LK	CS	NR	SE	LD	K	I	Y	O	T	S	N	FR	V	P	S	D	V	R	F	P	N	I	T	N	C	P	F	E	V																																
Civet SARS-CoV SZ3	KPLKLE	IN	IT	TER	LA	TA	PA	Q	D	TS	AA	AY	FV	G	L	K	EM	IR	D	ENG	T	DA	V	DC	SN	PL	AE	LK	CS	NR	SE	LD	K	I	Y	O	T	S	N	FR	V	P	S	D	V	R	F	P	N	I	T	N	C	P	F	E	V																																
Bat SL-CoV Rs3367	KPLKLE	IN	IT	TER	LA	TA	PA	Q	D	TS	AA	AY	FV	G	L	K	EM	IR	D	ENG	T	DA	V	DC	SN	PL	AE	LK	CS	NR	SE	LD	K	I	Y	O	T	S	N	FR	V	P	S	D	V	R	F	P	N	I	T	N	C	P	F	E	V																																
Bat SL-CoV RsSHC014	KPLKLE	IN	IT	TER	LA	TA	PA	Q	D	TS	AA	AY	FV	G	L	K	EM	IR	D	ENG	T	DA	V	DC	SN	PL	AE	LK	CS	NR	SE	LD	K	I	Y	O	T	S	N	FR	V	P	S	D	V	R	F	P	N	I	T	N	C	P	F	E	V																																
Bat SL-CoV Rs672	KPLKLE	IN	IT	TER	LA	TA	PA	Q	D	TS	AA	AY	FV	G	L	K	EM	IR	D	ENG	T	DA	V	DC	SN	PL	AE	LK	CS	NR	SE	LD	K	I	Y	O	T	S	N	FR	V	P	S	D	V	R	F	P	N	I	T	N	C	P	F	E	V																																
Bat SL-CoV Rf1	KPLKLE	IN	IT	TER	LA	TA	PA	Q	D	TS	AA	AY	FV	G	L	K	EM	IR	D	ENG	T	DA	V	DC	SN	PL	AE	LK	CS	NR	SE	LD	K	I	Y	O	T	S	N	FR	V	P	S	D	V	R	F	P	N	I	T	N	C	P	F	E	V																																
Bat SL-CoV Rm1	KPLKLE	IN	IT	TER	LA	TA	PA	Q	D	TS	AA	AY	FV	G	L	K	EM	IR	D	ENG	T	DA	V	DC	SN	PL	AE	LK	CS	NR	SE	LD	K	I	Y	O	T	S	N	FR	V	P	S	D	V	R	F	P	N	I	T	N	C	P	F	E	V																																
Bat SL-CoV Rp3	KPLKLE	IN	IT	TER	LA	TA	PA	Q	D	TS	AA	AY	FV	G	L	K	EM	IR	D	ENG	T	DA	V	DC	SN	PL	AE	LK	CS	NR	SE	LD	K	I	Y	O	T	S	N	FR	V	P	S	D	V	R	F	P	N	I	T	N	C	P	F	E	V																																
Bat SL-CoV HKU3-1	KPLKLE	IN	IT	TER	LA	TA	PA	Q	D	TS	AA	AY	FV	G	L	K	EM	IR	D	ENG	T	DA	V	DC	SN	PL	AE	LK	CS	NR	SE	LD	K	I	Y	O	T	S	N	FR	V	P	S	D	V	R	F	P	N	I	T	N	C	P	F	E	V																																
Bat SARS-related CoV BM48-31	KPLKLE	IN	IT	TER	LA	TA	PA	Q	D	TS	AA	AY	FV	G	L	K	EM	IR	D	ENG	T	DA	V	DC	SN	PL	AE	LK	CS	NR	SE	LD	K	I	Y	O	T	S	N	FR	V	P	S	D	V	R	F	P	N	I	T	N	C	P	F	E	V																																
Human SARS-CoV GZ02	FNATP	PS	Y	A	W	E	R	K	I	S	N	C	V	A	D	Y	S	V	M	N	-----	ST	FF	FC	Y	G	V	S	-----	SK	L	D	L	C	F	S	N	Y	A	D	S	F	V	V	R	K	D	V	R	Q	I	A	P	Q	T	G	V	I	A	D	N	Y	K	L	P	D	D	E	G	C	V	L	A	N	N	R	N	I	-----	DA	T	S	T	-----	NY	N	Y	N	K
Human SARS-CoV BJ01	FNATP	PS	Y	A	W	E	R	K	I	S	N	C	V	A	D	Y	S	V	M	N	-----	ST	FF	FC	Y	G	V	S	-----	SK	L	D	L	C	F	S	N	Y	A	D	S	F	V	V	R	K	D	V	R	Q	I	A	P	Q	T	G	V	I	A	D	N	Y	K	L	P	D	D	E	G	C	V	L	A	N	N	R	N	I	-----	DA	T	S	T	-----	NY	N	Y	N	K
Human SARS-CoV Tor2	FNATP	PS	Y	A	W	E	R	K	I	S	N	C	V	A	D	Y	S	V	M	N	-----	ST	FF	FC	Y	G	V	S	-----	SK	L	D	L	C	F	S	N	Y	A	D	S	F	V	V	R	K	D	V	R	Q	I	A	P	Q	T	G	V	I	A	D	N	Y	K	L	P	D	D	E	G	C	V	L	A	N	N	R	N	I	-----	DA	T	S	T	-----	NY	N	Y	N	K
Civet SARS-CoV SZ3	FNATP	PS	Y	A	W	E	R	K	I	S	N	C	V	A	D	Y	S	V	M	N	-----	ST	FF	FC	Y	G	V	S	-----	SK	L	D	L	C	F	S	N	Y	A	D	S	F	V	V	R	K	D	V	R	Q	I	A	P	Q	T	G	V	I	A	D	N	Y	K	L	P	D	D	E	G	C	V	L	A	N	N	R	N	I	-----	DA	T	S	T	-----	NY	N	Y	N	K
Bat SL-CoV Rs3367	FNATP	PS	Y	A	W	E	R	K	I	S	N	C	V	A	D	Y	S	V	M	N	-----	ST	FF	FC	Y	G	V	S	-----	SK	L	D	L	C	F	S	N	Y	A	D	S	F	V	V	R	K	D	V	R	Q	I	A	P	Q	T	G	V	I	A	D	N	Y	K	L	P	D	D	E	G	C	V	L	A	N	N	R	N	I	-----	DA	T	S	T	-----	NY	N	Y	N	K
Bat SL-CoV RsSHC014	FNATP	PS	Y	A	W	E	R	K	I	S	N	C	V	A	D	Y	S	V	M	N	-----	ST	FF	FC	Y	G	V	S	-----	SK	L	D	L	C	F	S	N	Y	A	D	S	F	V	V	R	K	D	V	R	Q	I	A	P	Q	T	G	V	I	A																														



Extended Data Figure 3 | Complete RdRP sequence phylogeny.

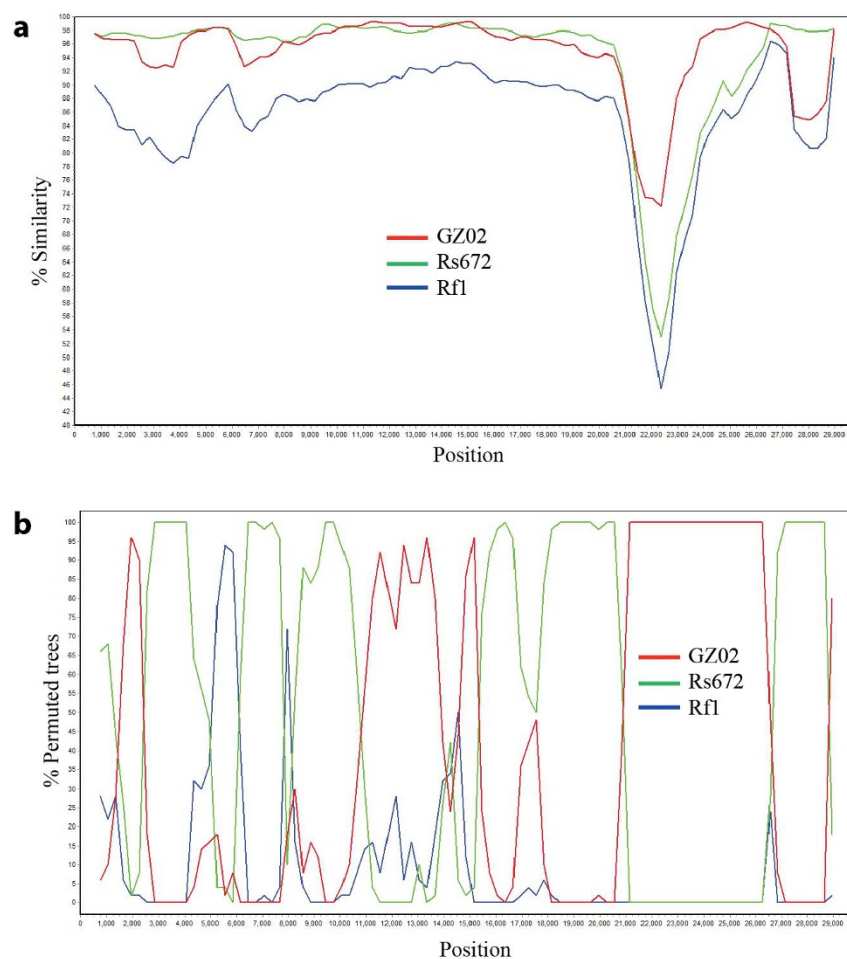
Phylogenetic tree of bat SL-CoVs and SARS-CoVs on the basis of complete RdRP sequences (2,796 nucleotides). Bat SL-CoVs RsSHC014 and Rs3367 are highlighted by filled circles. Three established coronavirus genera, *Alphacoronavirus*, *Betacoronavirus* and *Gammacoronavirus* are marked as α , β

and γ , respectively. Four CoV groups in the genus *Betacoronavirus* are indicated as A, B, C and D, respectively. MHV, murine hepatitis virus; PHEV, porcine haemagglutinating encephalomyelitis virus; PRCV, porcine respiratory coronavirus; FIPV, feline infectious peritonitis virus; IBV, infectious bronchitis coronavirus; BW, beluga whale coronavirus.



Extended Data Figure 4 | Sequence phylogeny of the complete S protein of SL-CoVs and SARS-CoV. Phylogenetic tree of bat SL-CoVs and SARS-CoVs on the basis of complete S protein sequences (1,256 amino acids).

Bat SL-CoVs RsSHC014 and Rs3367 are highlighted by filled circles. Bat CoV HKU9 was used as an outgroup.



Extended Data Figure 5 | Detection of potential recombination events.

a, b, Similarity plot (**a**) and bootscan analysis (**b**) detected three recombination breakpoints in the bat SL-CoV Rs3367 or SHC014 genome. The three breakpoints were located at the ORF1b (nt 20,827), M (nucleotides 26,553) and

N (nucleotides 28,685) genes, respectively. Both analyses were performed with an F84 distance model, a window size of 1,500 base pairs and a step size of 300 base pairs.

Extended Data Table 1 | Summary of sampling detail and CoV prevalence

Sampling time	Total number of swab or fecal samples collected	Number of CoV PCR positive samples (%)
April, 2011	14	1 (7.1)
October, 2011	10	3 (30)
May, 2012	54	4 (7.4)
September, 2012	39	19 (48.7)

Extended Data Table 2 | Genomic sequence identities of bat SL-CoVs with SARS-CoVs

Pairwise genomic nucleotide acids identity (%)												
CoVs	Genome size (nt)	Bat SARS-Like CoVs							Human and civet SARS-CoVs			
		SHC014	Rs672	Rp3	Rf1	Rm1	HKU3-1	BM48-31	GZ02	BJ01	Tor2	SZ3
3367	29,787	98.8	92.5	93.2	87.3	88.0	87.8	76.9	95.4	95.3	95.4	95.3
SHC014	29,787	-	92.6	93.2	87.3	88.1	87.8	77.0	95.2	95.1	95.1	95.1
Rs672	29,059		-	92.4	86.2	87.4	87.0	75.2	90.9	90.9	90.8	91.0
Rp3	29,736			-	88.3	90.3	89.6	77.0	92.1	92.0	92.1	92.0
Rf1	29,709				-	89.4	88.4	76.6	87.2	87.1	87.2	87.1
Rm1	29,749					-	90.1	76.4	87.6	87.5	87.5	87.5
HKU3-1	29,728						-	76.8	87.4	87.3	87.4	87.3
BM48-31	29,276							-	76.9	77.1	77.0	76.9
GZ02	29,760								-	99.6	99.6	99.7
BJ01	29,725									-	99.8	99.6
Tor2	29,751										-	99.5
SZ3	29,741											-

Extended Data Table 3 | Genomic annotation and comparison of bat SL-CoV Rs3367 with human/civet SARS-CoVs and other bat SL-CoVs

ORFs	Start-End (nt.)	No. of Nt.	No. of Aa.	TRS	ORF Identity nt/aa (%)										
					Human and civet SARS-CoVs				Bat SARS-like CoVs						
					GZ0 2	BJ01	Tor2	SZ3	Rs672	Rp3	Rf1	Rm1	HKU3-1	BM48-31	
P1a	265-13,398	13,134	4,377	ACGAAC ₁₉₂ AUG	96.7/97.9	96.6/97.9	96.8/97.9	96.8/98.1	93.3/94.2	95.5/96.9	88.1/94.0	87.9/93.3	87.9/94.2	76.3/80.8	
P1b	13,398-21,485	8,088	2,695		96.3/99.2	96.3/99.2	96.3/99.2	96.3/99.2	97.2/99.2	97.2/99.2	90.6/98.4	91.0/98.7	90.7/98.5	83.4/93.7	
S	21,492-25,262	3,771	1,256	ACGAACAUG	88.3/90.1	88.2/90.0	88.1/89.8	88.2/90.0	76.5/78.2	76.0/79.1	74.0/77.4	76.3/79.1	75.6/78.2	70.2/74.5	
(S1) *	21,493-23,535	2,043	681		78.2/81.1	78.2/80.9	78.1/80.6	78.2/81.1	65.1/62.2	63.9/63.0	62.9/62.5	64.7/63.3	65.2/63.4	62.2/64.7	
(S2) *	23,536-25,263	1,728	575		98.1/99.3	98.1/99.3	98.1/99.3	98.1/99.1	87.9/94.8	88.1/95.8	85.1/92.7	87.9/95.4	86/93.5	76.6/88.2	
ORF3a	25,271-26,095	825	274	ACGAAC ₂ AUG	99.2/98.1	98.6/97.0	98.7/97.0	98.5/96.7	90.4/90.8	84.1/84.3	88.8/86.8	83.6/84.3	83.1/82.4	72.1/71.2	
ORF3b	25,692-26,036	345	114		99.1/99.1	98.2/98.2	98.2/98.2	97.9/97.3	99.1/98.2	N/D	82.6/92.1	N/D	N/D	N/D	
E	26,120-26,350	231	76	ACGAAC ₂ AUG	98.7/98.6	98.7/98.6	98.7/98.6	98.7/98.6	99.1/98.6	97.8/98.6	96.5/96.0	96.1/97.3	97.4/98.6	91.3/93.4	
M	26,401-27,066	666	221	ACGAAC ₄₄ AUG	97.4/98.1	97.2/98.1	97.2/98.1	97.2/97.7	98.7/99.5	93.3/98.1	96.3/98.6	93.2/95.4	93.9/96.8	78.5/88.1	
ORF6	27,077-27,268	192	63	ACGAAC ₁₅₅ AUG	97.3/95.2	96.8/93.6	97.3/95.2	97.3/95.2	97.3/96.8	95.8/92.0	94.2/92.0	95.3/92	94.7/90.4	63.5/49.2	
ORF7a	27,276-27,644	369	122	ACGAACAUG	94.5/95.9	94.5/95.9	94.5/95.9	94.5/95.9	97.8/100	96.2/99.1	92.9/95.0	93.4/97.5	93.2/97.5	62.3/58.1	
ORF7b	27,641-27,776	135	44		96.2/93.1	96.2/93.1	96.2/93.1	96.2/93.1	99.2/100	99.2/100	97.7/97.7	99.2/100	93.3/95.4	62.9/63.6	
ORF8	27,782-28,147	366	121	ACGAACAUG	47.1/46.3	N/A	N/A	47.1/46.3	97.8/100	85.2/90.2	46.2/39.0	85.7/90.2	85.7/85.3	N/A	
N	28,162-29,430	1,269	422	ACGAAC ₈ AUG	98.3/99.5	98.4/99.5	98.4/99.5	98.4/99.5	98/98.5	96.6/97.6	93.7/95.2	96.2/97.1	95.9/96.2	77.9/87.2	
s2m	29,628-29,668	41			97.5	97.5	97.5	97.5	100	100	100	100	100	95.1	

*S1, the N-terminal domain of the coronavirus S protein responsible for receptor binding. S2, the S protein C-terminal domain responsible for membrane fusion.

The ORFs in the genome were predicted and potential protein sequences were translated. The pairwise comparisons were conducted for all ORFs at nucleotide acids (nt) and amino acids (aa) levels. The s2m were compared at nt level. TRS: Transcription regulating-sequences; N/D, not done; N/A, not available.

Extended Data Table 4 | Genomic annotation and comparison of bat SL-CoV RsSHC014 with human/civet SARS-CoVs and other bat SL-CoVs

ORFs	Start-End (nt.)	No. of Nt.	No. of Aa.	TRS	ORF Identity nt/aa (%)										
					Human and civet SARS-CoVs				Bat SARS-like CoVs						
					GZ0 ₂	BJ01	Tor2	SZ3	Rs672	Rp3	Rf1	Rm1	HKU3-1	BM48-31	
P1a	265-13,398	13,134	4,377	ACGAAC ₁₉₂ AUG	96.7/97.9	96.6/97.9	96.8/97.9	96.8/98.1	93.3/94.2	95.5/96.9	88.1/94.0	87.9/93.3	87.9/94.2	76.3/80.8	
P1b	13,398-21,485	8,088	2,695		96.3/99.2	96.3/99.2	96.3/99.2	96.3/99.2	97.2/99.2	97.2/99.2	90.6/98.4	91.0/98.7	90.7/98.5	83.4/93.7	
S	21,492-25,262	3,771	1,256	ACGAACAUG	88.3/90.1	88.2/90.0	88.1/89.8	88.2/90.0	76.5/78.2	76.0/79.1	74.0/77.4	76.3/79.1	75.6/78.2	70.2/74.5	
(S1) *	21,493-23,535	2,043	681		78.2/81.1	78.2/80.9	78.1/80.6	78.2/81.1	65.1/62.2	63.9/63.0	62.9/62.5	64.7/63.3	65.2/63.4	62.2/64.7	
(S2) *	23,536-25,263	1,728	575		98.1/99.3	98.1/99.3	98.1/99.3	98.1/99.1	87.9/94.8	88.1/95.8	85.1/92.7	87.9/95.4	86/93.5	76.6/88.2	
ORF3a	25,271-26,095	825	274	ACGAAC ₂ AUG	99.2/98.1	98.6/97.0	98.7/97.0	98.5/96.7	90.4/90.8	84.1/84.3	88.8/86.8	83.6/84.3	83.1/82.4	72.1/71.2	
ORF3b	25,692-26,036	345	114		99.1/99.1	98.2/98.2	98.2/98.2	97.9/97.3	99.1/98.2	N/D	82.6/92.1	N/D	N/D	N/D	
E	26,120-26,350	231	76	ACGAAC ₂ AUG	98.7/98.6	98.7/98.6	98.7/98.6	98.7/98.6	99.1/98.6	97.8/98.6	96.5/96.0	96.1/97.3	97.4/98.6	91.3/93.4	
M	26,401-27,066	666	221	ACGAAC ₄₄ AUG	97.4/98.1	97.2/98.1	97.2/98.1	97.2/97.7	98.7/99.5	93.3/98.1	96.3/98.6	93.2/95.4	93.9/96.8	78.5/88.1	
ORF6	27,077-27,268	192	63	ACGAAC ₁₅₅ AUG	97.3/95.2	96.8/93.6	97.3/95.2	97.3/95.2	97.3/96.8	95.8/92.0	94.2/92.0	95.3/92	94.7/90.4	63.5/49.2	
ORF7a	27,276-27,644	369	122	ACGAACAUG	94.5/95.9	94.5/95.9	94.5/95.9	94.5/95.9	97.8/100	96.2/99.1	92.9/95.0	93.4/97.5	93.2/97.5	62.3/58.1	
ORF7b	27,641-27,776	135	44		96.2/93.1	96.2/93.1	96.2/93.1	96.2/93.1	99.2/100	99.2/100	97.7/97.7	99.2/100	93.3/95.4	62.9/63.6	
ORF8	27,782-28,147	366	121	ACGAACAUG	47.1/46.3	N/A	N/A	47.1/46.3	97.8/100	85.2/90.2	46.2/39.0	85.7/90.2	85.7/85.3	N/A	
N	28,162-29,430	1,269	422	ACGAAC ₈ AUG	98.3/99.5	98.4/99.5	98.4/99.5	98.4/99.5	98/98.5	96.6/97.6	93.7/95.2	96.2/97.1	95.9/96.2	77.9/87.2	
s2m	29,628-29,668	41			97.5	97.5	97.5	97.5	100	100	100	100	100	95.1	

*S1, the N-terminal domain of the coronavirus S protein responsible for receptor binding. S2, the S protein C-terminal domain responsible for membrane fusion.

The ORFs in the genome were predicted and potential protein sequences were translated. The pairwise comparisons were conducted for all ORFs at nucleotide acids (nt) and amino acids (aa) levels. The s2m were compared at nt level. TRS: Transcription regulating-sequences; N/D, not done; N/A, not available.

Extended Data Table 5 | Cell lines used for virus isolation and susceptibility tests

Cell lines	Species (organ) origin	Medium	Infectivity
293T	Human (kidney)	DMEM+10%FBS	-
Hela	Human (cervix)		-
VeroE6	Monkey (kidney)		+
PK15	Pig (kidney)		+
BHK21	Hamster (kidney)		-
A549	Human (alveolar basal epithelial)		+
BK	<i>Myotis davidii</i> (kidney)	RPMI1640+10%FBS	-
RSKT	<i>Rhinolophus sinicus</i> (kidney)	DMEM/F12+10%FBS	+
MCKT	<i>Myotis chinensis</i> (kidney)		-
PaKi	<i>Pteropus alecto</i> (kidney)		-
RLK	<i>Rousettus leschenaulti</i> (kidney)		-

* Infectivity was determined by the presence of viral antigen detected by immunofluorescence assay.